Callery Homework 4 Output

Katie Callery

2/6/2018

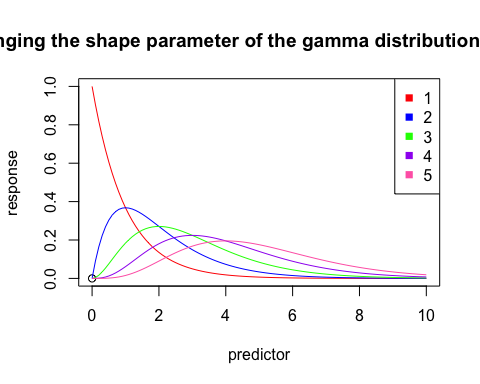
# Question 1

**Explore the properties of the gamma distribution using the gamma probability density function in R.** *Apply the curve function to explore how the shape and rate parameters alter the likelihood function.*

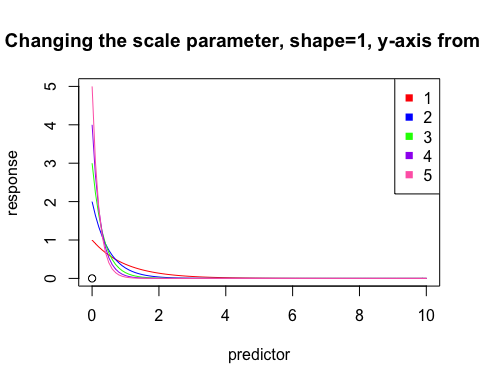
args(dgamma)

## function (x, shape, rate = 1, scale = 1/rate, log = FALSE)   
## NULL

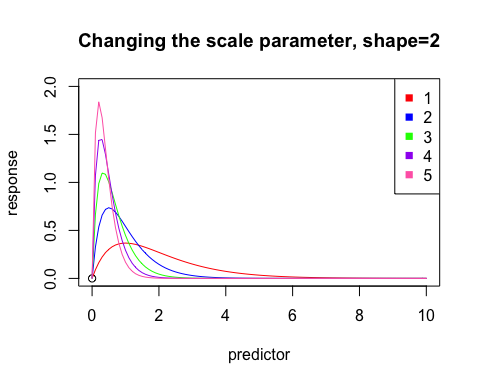
# function (x, shape, rate = 1, scale = 1/rate, log = FALSE)   
?dgamma  
# parameters are shape and scale, rate is an alternative way to specify the scale  
# the default is log = FALSE; I will keep the default  
  
# changing the shape parameter by increments of 1 from 1-5  
plot(0,0, xlim=c(0,10), ylim=c(0,1), xlab="predictor", ylab="response", main="Changing the shape parameter of the gamma distribution, scale=1")  
curve(dgamma(x, shape=1, rate = 1, log = FALSE), col="red", add=T)  
curve(dgamma(x, shape=2, rate = 1, log = FALSE), col="blue", add=T)  
curve(dgamma(x, shape=3, rate = 1, log = FALSE), col="green", add=T)  
curve(dgamma(x, shape=4, rate = 1, log = FALSE), col="purple", add=T)  
curve(dgamma(x, shape=5, rate = 1, log = FALSE), col="hotpink", add=T)  
legend("topright", legend=c("1","2","3","4","5"), col=c("red","blue","green","purple","hotpink"), pch=15)



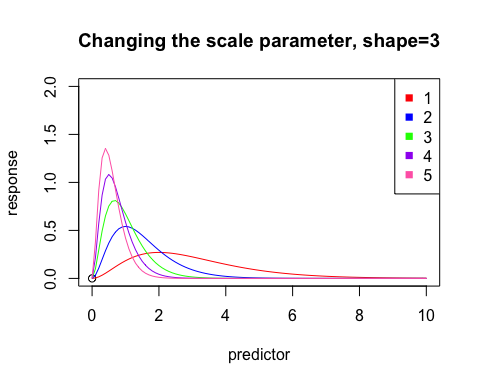
# changing the scale (rate) parameter by increments of 1 from 1-5  
plot(0,0, xlim=c(0,10), ylim=c(0,5), xlab="predictor", ylab="response", main="Changing the scale parameter, shape=1, y-axis from 0-5")  
curve(dgamma(x, shape=1, rate = 1, log = FALSE), col="red", add=T)  
curve(dgamma(x, shape=1, rate = 2, log = FALSE), col="blue", add=T)  
curve(dgamma(x, shape=1, rate = 3, log = FALSE), col="green", add=T)  
curve(dgamma(x, shape=1, rate = 4, log = FALSE), col="purple", add=T)  
curve(dgamma(x, shape=1, rate = 5, log = FALSE), col="hotpink", add=T)  
legend("topright", legend=c("1","2","3","4","5"), col=c("red","blue","green","purple","hotpink"), pch=15)



# changing the scale (rate) parameter by increments of 1 from 1-5  
plot(0,0, xlim=c(0,10), ylim=c(0,2), xlab="predictor", ylab="response", main="Changing the scale parameter, shape=2")  
curve(dgamma(x, shape=2, rate = 1, log = FALSE), col="red", add=T)  
curve(dgamma(x, shape=2, rate = 2, log = FALSE), col="blue", add=T)  
curve(dgamma(x, shape=2, rate = 3, log = FALSE), col="green", add=T)  
curve(dgamma(x, shape=2, rate = 4, log = FALSE), col="purple", add=T)  
curve(dgamma(x, shape=2, rate = 5, log = FALSE), col="hotpink", add=T)  
legend("topright", legend=c("1","2","3","4","5"), col=c("red","blue","green","purple","hotpink"), pch=15)



# changing the scale (rate) parameter by increments of 1 from 1-5  
plot(0,0, xlim=c(0,10), ylim=c(0,2), xlab="predictor", ylab="response", main="Changing the scale parameter, shape=3")  
curve(dgamma(x, shape=3, rate = 1, log = FALSE), col="red", add=T)  
curve(dgamma(x, shape=3, rate = 2, log = FALSE), col="blue", add=T)  
curve(dgamma(x, shape=3, rate = 3, log = FALSE), col="green", add=T)  
curve(dgamma(x, shape=3, rate = 4, log = FALSE), col="purple", add=T)  
curve(dgamma(x, shape=3, rate = 5, log = FALSE), col="hotpink", add=T)  
legend("topright", legend=c("1","2","3","4","5"), col=c("red","blue","green","purple","hotpink"), pch=15)

 (1) How do the shape and rate parameters determine the behavior of the function?

The shape parameter changes the shape of the curve in terms of at which x-value the slope = 0 (where the peak of the curve is).

As shape increases and scale remains static, the peak of the curve increases and seems to shift to shape-1 on the x-axis. Even when shape=1, the peak of the curve seems to be at shape-1, or 0 on the x-axis. This pattern seems conserved even as the scale changes, though the scale changes the shift of the peak by decreasing it by a factor of the scale. For example, when shape=2 and rate=1 (scale=1/rate), the curve peaks at 1 on the x-axis, or shape-1.

**When considering scale, it seems that the peak of the curve is at approximately (shape-1)/scale on the x-axis.When shape = 1, it appears that the curve peaks at y=rate, or at point (0,inverse scale). Increasing the scale parameter causes the peak to shift left on the x-axis, and shift up on the y-axis.**

1. Several different curves showing different parameter values for dgamma *(see above)*

# Question 2

**Find a published paper in your research area that uses either gamma or beta regression. Briefly (in one paragraph) answer the following questions.**

Hering et al. 2015. Contrasting the roles of section length and instream habitat enhancement for river restoration success: a field study of 20 European restoration projects. *Journal of Applied Ecology* 52:1518–1527.

*What were the response and predictor variables?*

The predictor variables were section length (“larger or smaller restoration projects”) and whether or not the section was actively restored.

The response variables were indicators of success in river restoration such as habitat diversity, aquatic organism abundance and diversity, and trophic connectivity between terrestial and aquatic habitats.

*How were the results presented?*

The authors presented their results as a series of boxplots depicting the Bray-Curtis dissimilarities in the percentages of the response variables for comparing both the restored and unrestored sections, and the long and short sections within the restored and unrestored categories.

From the boxplots, you could see which response variables were more affected by each of the predictor variables. Aquatic invertebrates (beetles) were relatively more abundant in the restored environments, and were the best indicator of restoration effort. You could also tell from the boxplots that the long and short sections of river did not differ in their percentages of the response variables, since the boxplots of the differences in these percentages overlapped zero for almost all of the variables.

*Do you agree with the interpretation of the results?*

I cannot say for certain, but I think so. It seems like the samples were independent enough from each other that they could say whether or not restoration effort or section length had impacts on river restoration results, and how large those impacts were.

However, I’m having a hard time understanding how the beta regression model played into this for these variables, and how it relates to Bray-Curtis, aside from the fact that they both handle values between 0-1. The Bray-Curtis dissimilarities seems straight-forward, but how is this different then a bunch of pairwise comparisons - akin to a t-test or an ANOVA?

# Question 3

**Apply beta or gamma regression to your own data (or a dataset from data dryad).** I’m going to use this data again, because I actually think the gamma distribution is better for this than the poisson distribution, having now learned about the gamma distribution. The level of carotenoids is not “countable” (does not come in regular intervals). So I will choose the **gamma distribution** for analyzing this data.

Sassani EC, Sevy C, Strasser EH, Anderson AM, Heath JA (2015) Plasma carotenoid concentrations of incubating American kestrels (Falco sparverius) show annual, seasonal, and individual variation and explain reproductive outcome. Biological Journal of the Linnean Society 117(3): 414-421. <https://doi.org/10.1111/bij.12653>

Sassani EC, Sevy C, Strasser EH, Anderson AM, Heath JA (2015) Data from: Plasma carotenoid concentrations of incubating American kestrels (Falco sparverius) show annual, seasonal, and individual variation and explain reproductive outcome. Dryad Digital Repository. <https://doi.org/10.5061/dryad.98c7s>

*Hypothesis: Male kestrels that arrive early to the breeding grounds are higher quality than those that arrive late.* *Prediction: Male kestrels caught earlier in the year will have higher carotenoid levels than those caught later in the year.*

carotenoid<-read.csv("ak\_carots\_ES2012\_proofed4.csv")  
# We are only going to look at carotenoid levels in male kestrels  
maleIndex<-which(carotenoid$sex == "male")  
carotenoidMale<-carotenoid[maleIndex,]  
attach(carotenoidMale)  
gamma\_model<-glm(carot~Jcap, family= Gamma(link="log"))  
# Why do you have to add the link=log part? What happens if you do not add that?  
exp(coef(gamma\_model))

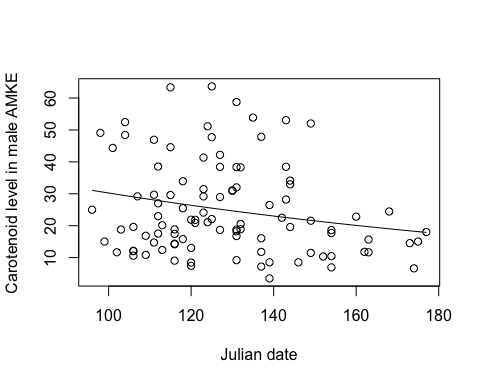
## (Intercept) Jcap   
## 59.8142795 0.9932041

#(Intercept) Jcap   
# 59.8142795 0.9932041   
exp(confint(gamma\_model))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 26.3977716 134.9427007  
## Jcap 0.9870516 0.9995272

# 2.5 % 97.5 %  
#(Intercept) 26.3977716 134.9427007  
#Jcap 0.9870516 0.9995272  
plot(Jcap, carot, xlab="Julian date", ylab="Carotenoid level in male AMKE")  
detach(carotenoidMale)  
curve(exp(4.091244421-0.006819098\*x), add=T)

 With every one unit increase of julian date, the level of carotenoids in captured male kestrels decreases by 0.7 %. At date 0 (Jan 1), the level of carotenoids is expected to be 59.8 (though kestrels are not captured that early; I could have centered around the mean, but Julian date as a number is informative).

Confidence intervals do not overlap 1, so the decrease is statistically significant.

# Question 4

**Use rgamma to simulate a gamma-distributed random variable that is related to a predictor variable.**

*Describe how you reparameterized R’s default gamma parameters to represent a mean value.*

I set the rate to be equal to shape/mu. Rate is equal to 1/scale. So scale would equal mu/shape in this case. Mu is the mean responses expected from the predictor data. This is incorporated into rgamma so that scale is equal to mu/shape. For a Gamma distribution, E(X) = a*s. The mean expected value could be plugged in here, so mu=a*s with a=shape and s=scale. Dividing mu (from our expontentiated linear equation) by the shape we set gives us the scale for our data, and then we can make a rgamma distribution for that data.

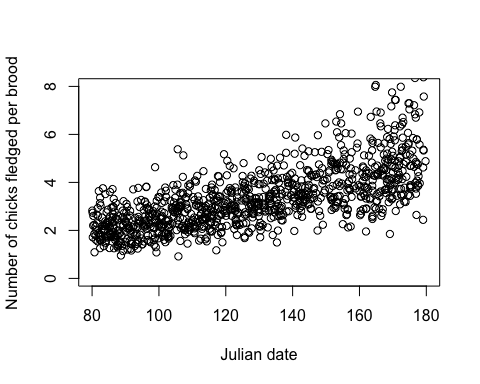
*Include a plot of your response vs. predictor variable, and a description of a biological process your data could represent.*

*Hint: y <- rgamma(n, rate = shape / y\_mean, shape = shape)*

sample\_size <- 1000   
  
predictor <- runif(sample\_size , min=80, max=180)  
  
Intercept <- 0  
  
Slope <- 0.009  
  
mu <- exp(Intercept + Slope \* predictor)   
  
shape <- 15  
  
response <- rgamma(sample\_size, rate=shape/mu, shape=shape)  
  
coef(glm(response~predictor,family=Gamma(link="log")))

## (Intercept) predictor   
## -0.03153529 0.00930766

#(Intercept) predictor   
# 0.45173995 0.02523849   
  
plot(predictor, response, ylim=c(0,8), xlim=c(80,180), xlab="Julian date", ylab="Number of chicks fledged per brood")

 I made these parameters up, but this may be an example from my data. Early in the breeding season, it is colder and there are less resources, so there may be higher nestling mortality, meaning fewer fledglings, regardless of clutch size. As the breeding seasons progresses, nestling mortality decreases, and the variety of clutch sizes is seen, from around 3-6 nestlings on average with some outliers.